WHAT IS CLAIMED IS:

1. A computer-implemented system for performing data mining applications, comprising:

(a) a computer having one or more data storage devices connected thereto, wherein a relational database is stored on one or more of the data storage devices;

(b) a relational database management system, executed by the computer, for accessing the relational database stored on the data storage devices; and

(c) an analytic application programming interface (API) that generates a set of scalable data mining functions, executed by the computer, for performing data mining operations directly within the database management system.

2. The system of claim 1 above, wherein the computer comprises a parallel processing computer comprised of a plurality of nodes, and each node executes one or more threads of the relational database management system to provide parallelism in the data mining operations.

3. The system of claim 1, wherein the scalable data mining functions process data collections stored in the relational database and produce results that are stored in the relational database.

4. The system of claim 1, wherein the scalable data mining functions are created by parameterizing and instantiating the analytic API.

5. The system of claim 1, wherein the scalable data mining functions comprise queries for execution by the relational database management system.

6. The system of claim 5, wherein the scalable data mining functions are dynamically generated queries comprised of combined phrases with substituting values therein based on parameters supplied to the analytic API.

7. The system of claim 6, wherein the scalable data mining functions are selected from a group of functions comprising Data Description functions, Data Derivation functions, Data Reduction functions, Data Reorganization functions, Data Sampling functions, and Data Partitioning functions.

8. The system of claim 7, wherein the Data Description functions comprise descriptive statistical functions.

9. The system of claim 7, wherein the Data Description functions are selected from a group comprising:

(1) descriptive statistics for one or more numeric columns, wherein the statistics are selected from a group comprising count, minimum, maximum, mean, standard deviation, standard mean error, variance, coefficient of variance, skewness, kurtosis, uncorrected sum of squares, corrected sum of squares, and quantiles,

(2) a count of values for a column,

(3) a calculated modality for a column,

(4) one or more bin numeric columns of counts with overlay and statistics options,

(5) one or more automatically sub-binned numeric columns giving additional counts and isolated frequently occurring individual values

(6) a computed frequency of one or more column values,

(7) a computed frequency of values for pairs of columns in a column list,

(8) a Pearson Product-Moment Correlation matrix,

(9) a Covariance matrix,

(10) a sum of squares and cross-products matrix, and

(11) a count of overlapping column values in one or more combinations of tables.

10. The system of claim 7, wherein the Data Derivation functions provide column derivations or transformations.

11. The system of claim 7, wherein the Data Description functions are selected from a group comprising:

(1) a derived binned numeric column wherein a new column is bin number,

(2) a n-valued categorical column dummy-coded into "n" 0/1 values,

(3) a n-valued categorical column recoded into n or less new values,

(4) one or more numeric columns scaled via range transformation,

(5) one or more columns scaled to a z-score that is a number of standard deviations from a mean,

(6) one or more numeric columns scaled via a sigmoidal transformation function,

(7) one or more numeric columns scaled via a base 10 logarithm function,

(8) one or more numeric columns scaled via a natural logarithm function,

(9) one or more numeric columns scaled via an exponential function,

(10) one or more numeric columns raised to a specified power,

(11) one or more numeric columns derived via user defined transformation function,

(12) one or more new columns derived by ranking one or more columns or expressions based on order,

(13) one or more new columns derived with quantile 0 to n-1 based on order and n,

(14) a cumulative sum of a value expression based on a sort expression,

(15) a moving average of a value expression based on a width and order,

(16) a moving sum of a value expression based on a width and order,

(17) a moving difference of a value expression based on a width and order,

(18) a moving linear regression value derived from an expression, width, and order,

(19) a multiple account/product ownership bitmap,

(20) a product ownership bitmap over multiple time periods,

(21) one or more counts, amount, percentage means and intensities derived from a transaction summary,

(22) one or more variabilities derived from transaction summary data,

(23) one or more derived trigonometric values and their inverses, including sin, arcsin, cos, arccos, csc, arccsc, sec, arcsec, tan, arctan, cot, and arccot, and

(24) one or more derived hyperbolic values and their inverses, including sinh, arcsinh, cosh, arccosh, csch, arccsch, sech, arcsech, tanh, arctanh, coth, and arccoth.

12. The system of claim 7, wherein the Data Reduction functions provide matrix building operations to reduce the amount of data required for analytic algorithms.

13. The system of claim 7, wherein the Data Reduction functions are selected from a group comprising:

(1) build one or more data reduction matrices from a group comprising: (i) a Pearson-Product Moment Correlations matrix; (ii) a Covariances matrix; and (iii) a Sum of Squares and Cross Products (SSCP) matrix,

(2) export a resultant matrix, and

(3) restart a matrix operation.

14. The system of claim 7, wherein the Data Reorganization functions provide an ability to reorganize data by joining or de-normalizing pre-processed results into a wide analytic data set.

15. The system of claim 7, wherein the Data Reorganization functions are selected from a group comprising:

(1) create a de-normalized new table by removing one or more key columns, and

(2) join a plurality of tables or views into a combined result table.

16. The system of claim 7, wherein the Data Sampling function provides an ability to construct a new table containing a randomly selected subset of the rows in an existing table or view.

17. The system of claim 7, wherein the Data Sample function selects one or more data samples of specified sizes from a table.

18. The system of claim 7, wherein the Data Partitioning function provides an ability to construct a new table containing at least one randomly selected subset of the rows in an existing table or view, wherein the subsets are mutually distinct but all-inclusive subsets of data.

19. The system of claim 7, wherein the Data Partitioning function selects one or more data partitions from a table using a database internal hashing technique.

20. The system of claim 1, wherein results of the data mining operations are stored in the relational databases.

21. The system of claim 1, wherein the relational database management system further comprises an analytical logical data model that stores metadata and processing results from the Scalable Data Mining Functions.

22. A method for performing data mining applications, comprising:

(a) storing a relational database on one or more data storage devices connected to a computer;

(b) accessing the relational database stored on the data storage devices using a relational database management system; and

(c) utilizing a comprehensive set of parameterized analytic capabilities for performing data mining operations directly within a massively parallel relational database management system.

23. An article of manufacture comprising logic embodying a method for performing data mining applications, comprising:

(a) storing a relational database on one or more data storage devices connected to a computer;

(b) accessing the relational database stored on the data storage devices using a relational database management system; and

(c) utilizing a comprehensive set of parameterized analytic capabilities for performing data mining operations directly within a massively parallel relational database management system.

B2 Add

add a1